Die Kunst, Schachwettkampfresultate zu bewerten

Fragestellung:
Welche Schlüsse kann man aus Wettkampfresultaten ziehen?
Wieviel Patien müssen gespielt werden, um ein „gerechtes Resultat" zu erhalten?
Die Statistik und die Wahrscheinlichkeitstheorie können uns helfen, diese Fragen zu beantworten.

Ausgangspunkt
Wir haben ein Match-Resultat zweier Schachcomputer. Wir nehmen an, daß der Computer C mit 30-20 gegen Computer D gewonnen hat.Die Partien bezeichnen wir mit x1, x2, .. x50 und die Resultate des Computers C können mit 0, 1/2 resp. 1 bewertet werden. Nun wollen wir Computer C's Resultat näher untersuchen. Mit m bezeichnen wir die Punktzahl geteilt durch die Anzahl der gespielten Partien. Für unseren Wettkampf ist also m = 30/50 0.6 (Der Wert von m muß zwischen 0 und 1 liegen). Mit m versuchen wir auszudrücken, daß C im Durchschnitt mit 0,6 Punkten gegen D gewinnt. Dieses ist aber eine allzu kategorische Behauptung, da das Wettkampfresultat etwas unsicher sein kann.

Intervall
Eine bessere Aussage wäre: C gewinnt über D im Durchschnitt mit 0,6 +− A Punkten. Nun haben wir ein Intervall, dessen Breite von A bestimmt wird. A kann umso kleiner angenommen werden, je sicherer das Resultat ist.
Die Anzahl der gespielten Partien bestimmt die Sicherheit der Aussage m. Je mehr Partien gespielt werden, desto sicherer ist das Resultat und desto kleiner kann A sein. Hier folgt nun eine Formel zur Berechnung von A.

$$A = 1.96\, s\sqrt{n}$$

Die Konstante 1.96 wird später erklärt, n ist die Anzahl der gespielten Partien und s ist die Standardabweichung, die wie folgt berechnet wird:

$$s = \sqrt{\frac{(x1 - m)^2 + (x2 - m)^2 + \ldots (xn - m)^2}{n - 1}}$$

Um s berechnen zu können, muß man wissen, wieviele Partien gewonnen, verloren wurden resp. ein Remis ergaben. Für unser Beispiel nehmen wir an, daß C 28 Partien gewann, 18 verlor und 4 Partien ein Remis erbrachten. Dann erhalten wir:

$$s = \sqrt{\frac{28(1-0.6)^2 + 4(0.5-0.6)^2 + 18(0-0.6)^2}{49}} \approx 0.47$$

und somit: $A = 1.96 \times .47\sqrt{50} \approx 0.13$

Ein erster Versuch, unser Wettkampfresultat zu deuten
C erhält im Durchschnitt pro Partie 0.6 +− 0.13 Punkte (gegenüber D) und das Intervall liegt zwischen 0.47 und 0.73 Punkten. In Prozenten ausgedrückt (mit 100 multipliziert) wären es 47% - 73%.
Wir könnten auch mit 50 multiplizieren (= Anzahl der gespielten Partien) und dann sagen: C dürfte mit 23.5 - 36.5 Punkten gewinnen, wenn der Wettkampf wiederholt wird. Diese verschiedenen Zahlenspiele drücken dasselbe nur mit verschiedenen Ziffern aus.

Sicherheit der Aussage
Sind nun unsere Schlüsse völlig richtig? Nein, leider nicht. Glücklicherweise gibt es aber eine Möglichkeit, auch hierüber Klarheit zu gewinnen. Die Konstante 1.96 in der Formal für A ergibt eine 95%-ige Sicherheit der Aussage. Dieses bedeutet, daß unsere Prophezeiung bei 100 gespielten Partien 5 mal falsch sein wird. Durch Abänderung der Konstanten 1.96 können wir verschiedene Grade der „Sicherheit" erhalten:
Sicherheit: 80% - Konstante 1.28; 90% - 1.64; 95% - 1.96; 99% - 2.58; 99,5% - 2.81; 99,9% - 3.29.

Eine bessere Deutung des Wettkampfresultates
Zu unserem Beispiel, in dem C mit 30 gegen 20 Punkten über D gewann, können wir nun folgendes sagen: Mit 95% Sicherheit gewinnt C im Durchschnitt pro Partie zwischen 0.47 und 0.73 Punkten. Wir sehen hier, daß man mit 95% Sicherheit nicht behaupten kann, daß C besser ist als D. Um dieses sagen zu können, müßte die unterste Grenze größer sein als 0.5. Einen anderen Grad der Sicherheit erhält man durch Austausch der Konstanten 1.96 durch eine andere (siehe Tabelle). Für eine 99,5% Sicherheit nehmen wir also 2.81 und erhalten

$$A = 2.81 \times 0.47\sqrt{50} \approx 0.19$$

Dann kann man sagen, daß C mit 99,5% Sicherheit im Durchschnitt mit 0.41 bis 0.79 Punkten gegen D gewinnt.

Vorbehalt
Wir haben mit noch einem Unsicherheitsfaktor zu rechnen. Alles bis jetzt Gesagte gilt nur unter der Voraussetzung, daß die Resultate einer Normalverteilung folgen. Dieses wird aber in der Wirklichkeit nicht der Fall sein, sondern die Resultate entsprechen nur einer „ungefähren" Normalverteilung. Diese Tatsache kann das Resultat manchmal recht beachtlich beeinflussen. So kann man zum Beispiel eine negative untere Grenze erhalten, resp. eine obere, die größer ist als 1. Dieses ist natürlich unmöglich, und in solchen Fällen rundet man auf 0 resp. 1 ab. Am besten aber ist, wenn man mehr Partien spielt, denn je mehr Patien gespielt werden, desto mehr nähert man sich der Normalverteilung. Eine große Anzahl gespielter Partien haben einen guten Einfluß auf A, das ja das Intervall bestimmt, und wir in der Formel für A mit der Quadratwurzel der Anzahl der gespielten Partien dividieren..
Eine richtigere Ausdrucksweise wäre somit: Mit ungefähr 95% Sicherheit .. usw.

Wieviel Partien sollten gespielt werden?
Die Theorie sagt „eine große Anzahl", teilweise um der Normalverteilung näher zu kommen, teilweise um einen guten Wert für s zu erhalten. Verschiedene Gelehrte geben nun verschiedene Werte dafür an, was unter „einer großen Anzahl" zu verstehen ist. Keiner (soweit mir bekannt ist) will in unserem Falle unter 30 Partien gehen. Einige nennen 50 Partien. Wüßten wir, daß m genau der Normalverteilung entspricht oder wäre uns der genaue Wert von s bekannt, so würden wohl 20 Partien genügen. Wissen wir, daß wenigstens 1 Partie ein Remis ergibt und spielt man eine unendliche Anzahl von Partien, so wissen wir das „s" höchstens 0.5 ergibt. In diesem Falle setzen wir s = 0.5 in die Formal für A ein. Gewöhnlich erhalten wir dann ein gößeres Intervall, aber dagegen können wir eine geringere Anzahl von Partien spielen (wenigstens 20). Diesen Kunstgriff können wir auch anwenden, wenn uns die einzelnen Resultate nicht bekannt sind- .Wir nehmen also an: s = 0.5. Im übrigen kann gesagt werden, daß das Intervall kleiner wird je mehr Partien ein Remis ergeben und daß diese Methode nicht angewendet werden kann, wenn alle Partien dasselbe Resultat ergeben.

Die Rangliste
Aus dem Gesagten dürfte hervorgehen, daß eine Rangliste, die sich von 10 bis 20 Partien herleitet, sehr ungenau sein kann und sicherlich auch meistens ist. Auch eine Rangliste, die sich auf 50 Partien bezieht, muß man skeptisch betrachten. Man muß bedenken, daß es sich bei einer Rangliste um eine exakte Zahl (Wert) handelt und nicht um ein Intervall, und dieses ergibt dann große Fehler. Man sollte sich also über Fehler von +− 50 Punkten in einer Rangliste nicht wundern, wenn auch mehr als 50 Partien gespielt worden sind.

Wettkampfresultat
Dieselbe Methode kann bei Wettkämpfen angewendet werden (Jeder spielt gegen jeden). Multipliziert man die Intervallgrenze mit der Anzahl Runden, so erhält man ein Intervall mit einem wahrscheinlichen Punktresultat für die Spieler, wenn der Wettkampf wiederholt wird.

Mensch und Schachcomputer
Eine Voraussetzung dafür, daß diese Methode richtige Resultate ergibt, ist, daß jede Partie nicht von anderen Partien beeinflußt wird. Dies dürfte beim Spiel Computer gegen Computer gegeben sein, nicht aber, wenn ein Mensch am Spiele aktiv teilnimmt. Denn wir wissen, wie sehr ein Resultat die menschlichen Sinne beeinflussenn kann und das weitere Spiel beeinflußt. Hier haben wir also eine neue Fehlerquelle beim Spiel Mensch gegen Mensch oder Mensch gegen Computer.

Abschluß
Es ist eine gute Regel, sich im voraus für ein Intervall zu entscheiden und sich dann konsequent daran zu halten. Man ändert also nicht die Voraussetzungen, wenn verschiedene Resultate beurteilt werden sollen.
Der Statistiker wird gemerkt haben, daß hier weder Fachausdrücke noch Symbole der Statistik angewandt wurden. Der Gedanke war, hierdurch den Text mehr gemeinverständlich zu machen. Jean-Peter Fendrich/Schweden

Jean-Peter Fendrich (Schweden): Die Kunst, Schachwettkampfresultate zu bewerten

(Quelle: https://rochadeeuropa.com/ Nr. 226 – Mai 1983) (photo copyright © by http://www.schaakcomputers.nl/) (600 dpi)

First Published on February 2, 2018 – Size: 7.843 KB